

AMENDMENT TO THE CLAIMS

1-26. (Cancelled)

27. (Currently Amended) A method for evaluating a word segmentation language model, comprising:

building the word segmentation language model based on an annotated corpus;
applying the language model to a test corpus of unsegmented text different from
the annotated corpus to provide an output indicative of words in the test
corpus and a word type indication for each word, the word type indication
being one of a plurality of word type indications;

comparing the word type indication for each word in the output of the language
model with predefined word type indications of words of the test corpus;
and

generating a quantitative value that represents a level of precision with which
word type indications were applied in the output indicative of words in the
test corpus, wherein generating comprises generating based on a
comparison of the word type indication for words in the output to the
predefined word type indications.

~~evaluating the language model based on the comparison of the word type
indication for each word in the output and the predefined word type
indications to provide an indication of effectiveness of the language model
as a function of the word type indications identified by the language
model.~~

28-30. (Cancelled)

31. (Previously Presented) A method of evaluating word segmentation models, comprising:

- using a first word segmentation model to segment a corpus of text into words and apply tags to the words indicative of one of a plurality of word types, the words and tags forming a first output;
- using a second word segmentation model to segment the corpus of text into words and apply tags to the words indicative of one of the plurality of word types, the words and tags forming a second output;
- comparing the first output to a predefined indication of words and tags of the words indicative of one of the plurality of word types from the corpus of text to provide a first set of values for each of the plurality of word types indicative of how the first word segmentation model recognizes each of the plurality of word types;
- comparing the second output to the predefined indication of words and tags of the words indicative of one of the plurality of word types from the corpus of text to provide a second set of values for each of the plurality of word types indicative of how the second word segmentation model recognizes each of the plurality of word types; and
- comparing the first set of values and the second set of values to determine effectiveness of the first word segmentation model and the second word segmentation model with respect to each of the plurality of word types.

32. (Previously Presented) The method of claim 31 wherein the first set of values is based on matches between the first output and the predefined indication and wherein the second set of values is based on matches between the second output and the predefined indication.

33. (Cancel)

34. (New) The method of claim 27 wherein generating the quantitative value comprises generating a quantitative value based on a comparison of word type indications of words in the output that match predefined word type indications assigned to the same words in the test corpus.

35. (New) The method of claim 27 wherein generating the quantitative value comprises generating a quantitative value that is indicative of how frequently a word type indication in the output matches a corresponding predefined word type indication in the test corpus.

36. (New) The method of claim 27 wherein generating the quantitative value comprises generating a quantitative value that is indicative of how frequently a word type indication, assigned to a word in the output, matches a predefined word type indication assigned to a same word in the test corpus.

37. (New) The method of claim 27 wherein generating a quantitative value further comprises generating a quantitative value that represents a level of precision with which person name word type indications were applied in the output.

38. (New) The method of claim 27 wherein generating a quantitative value further comprises generating a quantitative value that represents a level of precision with which location name type indications were applied in the output.

39. (New) The method of claim 27 wherein generating a quantitative value further comprises generating a quantitative value that represents a level of precision with which organization name word type indications were applied in the output.

40. (New) The method of claim 27 wherein generating a quantitative value further comprises generating a quantitative value that represents a level of precision with which overlapping ambiguous string word type indications were applied in the output.

41. (New) The method of claim 27 wherein generating a quantitative value further comprises generating a quantitative value that represents a level of precision with which covering ambiguous string word type indications were applied in the output.
42. (New) The method of claim 31 wherein comparing to provide a first set of values for each of the plurality of word types comprises comparing to provide a first set of values for a person name word type.
43. (New) The method of claim 31 wherein comparing to provide a first set of values for each of the plurality of word types comprises comparing to provide a first set of values for a location name word type.
44. (New) The method of claim 31 wherein comparing to provide a first set of values for each of the plurality of word types comprises comparing to provide a first set of values for an organization name word type.
45. (New) The method of claim 31 wherein comparing to provide a first set of values for each of the plurality of word types comprises comparing to provide a first set of values for a overlapping ambiguous string word type.
46. (New) The method of claim 31 wherein comparing to provide a first set of values for each of the plurality of word types comprises comparing to provide a first set of values for a covering ambiguous string word type.